

Были рассмотрены четыре текста: два текста, содержащие лексему «конструкция» со значением «pattern (в языке, например)» (L1, L2) и два текста, содержащие лексему «конструкция» со значением «строительная конструкция» (B1, B2). Для каждого текста были выделены **значимые леммы**. После каждого текста указано количество выделенных значимых лемм.

L1. Английский **язык** содержит свыше десятка различных **вопросительных слов**, с помощью которых и составляются такие **конструкции**. Употребляются они со всеми возможными разновидностями прошедшего **времени**, и запомнить их построение очень просто: **вопросительные** слова всегда стоят в начале **предложения**. Иными словами, необходимо создать любой уточняющий вопрос и добавить вперед соответствующую случаю **конструкцию**. N = 6

L2. Русский **язык** предоставляет **говорящему** и пишущему широкие возможности выбора между **синонимичными** (близкими по **смыслу**) **грамматическими** **конструкциями**. Часто от того, какой вариант предпочтет автор документа, зависит степень понятности и убедительности сообщения. Выбор **пассивной конструкции** придает **тексту** большую официальность. Подчеркивается сам факт осуществления действия вне зависимости от того, кто его произвел. N=7

B1. Главные составляющие **несущей конструкции крыши** - **стропильные фермы** и **обрешетка**. **Кровля** - всего лишь наружная часть крыши, которая укладывается на несущую конструкцию, состоящую из стропильных ферм и обрешетки. В основе стропильной фермы лежит треугольник, как наиболее жесткая и экономичная конструкция. N=5

B2. **Сборные строительные конструкции** на **электростанциях** в малых объемах применяются много лет. При изготовлении сборных строительных **конструкций** на **домостроительных комбинатах** или на **заводах железобетонных изделий** либо при **возведении зданий** из кирпича должны быть оставлены **отверстия** для прохода **труб** в **стенах, перекрытиях** и **перегородках**, а также борозды в стенах здания для скрытой прокладки **трубопроводов**. N=16

По формуле индекса Дайса $I_d = 2 * X \& Y / (X + Y)$, где I_d -- индекс Дайса, X -- множество значимых лемм для первого сравниваемого текста, Y -- множество значимых лемм для второго сравниваемого текста, были вычислены значения индекса Дайса для возможных пар:

	L1	L2	B1	B2
L1	1			
L2	$2/(6+7) \approx 0,15$	1		
B1	$0/11 = 0$	$0/12 = 0$	1	
B2	$0/22 = 0$	$0/12 = 0$	$0/21 = 0$	1

Как легко заметить, среди всех возможных пар различающихся текстов только для L1 и L2 индекс Дайса отличается от нуля.